ISSN 0104-0499

PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

DEs-UFSCar e SME-ICMC-USP

A FLEXIBLE BAYESIAN PARTITION MODELLING FOR LONG-TERM SURVIVAL DATA

Vera. L. D. Tomazella
Jhon F. B. Gonzales
Mário de Castro

RELATÓRIO TÉCNICO

TEORIA E MÉTODO – SÉRIE A

Setembro/2013
nº 257

A flexible Bayesian partition modelling for long-term survival data

Vera. L. D. Tomazella[†], Jhon F.B. Gonzales [*†] and Mário de Castro[‡]

[†]Departamento de Estatística, Universidade Federal de São Carlos, Rod. Washington Luíz, km 235, 13565-905, São Carlos, SP, Brazil
[‡]Instituto de Ciências Matemáticas e da Computação, Universidade de São Paulo, Avenida Trabalhador São-carlense, 400 - Centro, 13566-590, São Carlos, SP, Brazil

September 30, 2013

Abstract

In this paper we propose a Bayesian partition modeling for lifetime data in presence of a cure fraction by considering a local structure generated by a tessellation which depends on covariates. In this modelling we including information of nominal qualitative variables with more than two categories or ordinal qualitative variables. The proposed modeling is based on a promotion time cure model structure but assuming that the number of competing causes follow a power series distribution. It is an alternative modeling strategy to the conventional survival regression modeling generally used for modeling lifetime data in presence of a cure fraction, which models the cure fraction through a (generalized) linear model of the covariates. An advantage of our approach is its ability to capture the effects of covariates in a local structure. The flexibility of having a local structure is crucial to capture local effects and features of the data. The modelling is illustrated on two real melanoma data sets.
Keywords: Survival Analysis; Long-term survival models; Bayesian partition.

1 Introduction

With rapid progress in the medical and health sciences, many datasets dealing with time to relapse now reveal a substantial proportion of patients who are expected non-susceptible to the occurrence to event interest (i.e. who

[*]jhonbg@gmail.com

are cured). Lifetime data in which there are sampling units non-susceptible to the occurrence of the event of interest, which can usually be caused by different latent competing causes, are common in applications from various areas, such as medical, financial and industrial ones.

The competing causes are latent in the sense that there is no information about which factor was responsible for the component failure (or individual death). The statistical literature for modeling lifetime data in presence of a cure fraction and latent competing causes is by now vast and growing rapidly. Interested readers can refer to Boag (1949), Berkson & Gage (1952), Maller & Zhou (1996), Yakovlev & Tsodikov (1996), Chen *et al.* (1999), Tsodikov *et al.* (2003), Yin & Ibrahim (2005), Cooner *et al.* (2007), de Castro *et al.* (2009), Rodrigues *et al.* (2009b) and Cancho *et al.* (2011) among others.

Usually the cure fraction models are modifications of the mixture cure model (Berkson & Gage, 1952) or the promotion time model (Yakovlev & Tsodikov, 1996; Chen *et al.*, 1999). Recent papers suggest other flexible models for data with a cure fraction as Cooner *et al.* (2007), Gu *et al.* (2011) and Rodrigues *et al.* (2011), where it is assumed that the number of latent competing causes is a random variable with a discrete distribution such as the binomial, Poisson, geometric and so on.

The conventional approaches described above model the cure fraction through a (generalized) linear model of the covariates. However, there are situations where the use of this approach suffers from poor predictive results (Hoggart & Griffin, 2001). For extra flexibility, in this paper, we model the covariate effect locally in the cure fraction using the so called Bayesian partition model (BPM) proposed by Holmes *et al.* (1999). Instead of considering a link function to connect the cure fraction to covariates, the BPM uses a local structure generated by a tessellation and which depends on covariates.

The partition modeling is not a new idea and various authors have studied partition models from a Bayesian viewpoint. Among them we can mention Yao (1984), Green & Sibson (1978) and Heikkinen (1998), with applications in various areas, such as geostatistics, genetics and finance, among others.

However, in medical and epidemiological studies, often interest focuses on studying nominal qualitative variables with more than two categories or ordinal qualitative variables. For example, researchers may be interested on study of cancer of melanoma, where we have important factors as disease stage, tumor size, category node, among other. So the aim of the present paper is to present flexible methodologies suited to incorporate information on nominal qualitative variables with more than two categories or ordinal qualitative variables into Bayesian analysis. However, to the best of our knowledge the present paper is the first attempt to consider partition modelling in the context of lifetime data in presence of competing risks and a cure fraction.

For lifetime data with a cure fraction, the partition model schemes usually

split the time axis as can be seen Ibrahim *et al.* (2001a) and Kim *et al.* (2007), but these models also make use of a link function to connect the cure fraction to covariates.

In this paper however, we propose a BPM for lifetime data in presence of a cure fraction by considering a local structure generated by a tessellation and which depends on covariates. The proposed modeling is based on a promotion time cure rate model structure, but assuming that the number of competing causes follow a power series distribution. An advantage of our approach is its ability to capture the effects of covariates in a local structure. The flexibility of having a local structure is crucial to capture local effects and features of the data.

The paper is organized as follows. In Section 2 we present the cure rate model with latent competing causes structure. In Section 3 we present the BPM scheme considered as well as the inferential approach. In Section 4 we will apply the proposed BPM to a real dataset on melanoma data. The paper is ended with Section 5 where we present some final comments.

2 Model formulation

Let N be a discrete random variable representing the latent number of competing causes needed to the occurrence of a particular event of interest. We assume that N has the power series distribution (Johnson *et al.*, 2005) with probability mass function

$$P[N = k; \theta] = \frac{a_k \theta^k}{\eta(\theta)}, \quad k = 0, 1, 2, \ldots, \quad \theta > 0, \tag{1}$$

where $a_k > 0$ and $\eta(\theta) = \sum_{k=0}^{\infty} a_k \theta^k < +\infty$. In (1), θ and $\eta(.)$ are called the power parameter and the series function, respectively. The probability generating function of N is given by

$$G(s) = \frac{\eta(\theta s)}{\eta(\theta)}, \quad 0 \leq s \leq 1. \tag{2}$$

For different series functions $\eta(.)$, Kosambi (1949) and Noack (1950) obtained the following results we present in Table 1. The parameters K and τ in the Table 1 are non-negative integer numbers. Moreover, the support of logarithmic distribution was shifted to $N = 0, 1, 2, \ldots$.

Conditioned on N, let Z_v, $v = 1, \ldots, N$ be i.i.d. random variables with cumulative distribution function $F(t)$ and survival function $S(t) = 1 - F(t)$, where Z_v is the time of occurrence of an particular event of interest due to the v-th cause. For instance, in a biological scenario N may denote the number of carcinogenic cells which can produce a detectable tumor (Yakovlev

Table 1: Distribution of N for different series function $\eta(.)$

$\eta(\theta)$	Θ	Support	a_k	a_0	Distribution $p(N\|\theta)$
$(1+\theta)^K$	$\theta > 0$	$k = 0, 1, \ldots, K$	$\binom{K}{k}$	1	Binomial Bi $\left(K, \frac{\theta}{1+\theta}\right)$
e^θ	$\theta > 0$	$k = 0, 1, 2, \ldots$	$\frac{1}{k!}$	1	Poisson Poi(θ)
$(1-\theta)^{-\tau}$	$\theta \in (0,1)$	$k = 0, 1, 2, \ldots$	$\binom{\tau+k-1}{\tau-1}$	1	Negative binomial Bn(τ, θ)
$\frac{-\log(1-\theta)}{\theta}$	$\theta \in (0,1)$	$k = 0, 1, 2, \ldots$	$\frac{1}{(k+1)}$	1	Logarithmic Lg(θ)

& Tsodikov, 1996). The observable time of occurrence of event of interest is defined by

$$T = \min\{Z_1 \ldots, Z_N\}. \tag{3}$$

Under this setup, according to Tsodikov *et al.* (2003), Rodrigues *et al.* (2009a), among others, the survival function for the population is given by

$$S_{pop}(t) = G(S(t)) = \frac{\eta(\theta S(t))}{\eta(\theta)}. \tag{4}$$

The survival function $S_{pop}(t)$ given in (4) is not a proper survival function by the fact that

$$p_0 = \lim_{t\to\infty} S_{pop}(t) = \frac{\eta(\theta S(\infty))}{\eta(\theta)} = \frac{a_0}{\eta(\theta)} < 1, \tag{5}$$

where p_0 denotes the proportion of cured that may be present in the population from which the data were taken. So, the improper density and risk functions associated with long-term survival function in (4) are given respectively by

$$f_{pop}(t) = \frac{\eta'(\theta S(t))}{\eta(\theta)}\theta f(t) \text{ and } h_{pop}(t) = \frac{\eta'(\theta S(t))}{\eta(\theta S(t))}\theta f(t), \tag{6}$$

where $f(t)$ denotes the (proper) density function of the lifetime Z and $\eta'(\theta S(t)) = \mathrm{d}\eta(s)/\mathrm{d}s|_{s=\theta S(t)}$.

In the Table 2 show the improper survival and density functions for different distributions of N. We changed the parametrization in the binomial distribution for considering $\theta^* = \theta/(1+\theta)$.

Table 2: Survival function $S_{pop}(t)$, density function $f_{pop}(t)$ and cure fraction p_0 for different distributions for N.

$p(N\|\theta)$	$S_{pop}(t)$	$f_{pop}(t)$	p_0
Bi(K,θ^*)	$(1-\theta^*+\theta^* S(t))^K$	$K\theta^* f(t)(1-\theta^*+\theta^* S(t))^{K-1}$	$(1-\theta^*)^K$
Poi(θ)	$\exp(-\theta F(t))$	$\theta f(t)\exp(-\theta F(t))$	$e^{-\theta}$
Bn(τ,θ)	$\left(\frac{1-\theta}{1-\theta S(t)}\right)^\tau$	$\frac{\tau\theta(1-\theta)^\tau f(t)}{(1-\theta S(t))^{\tau+1}}$	$(1-\theta)^\tau$
Lg(θ)	$\frac{\log(1-\theta S(t))}{S(t)\log(1-\theta)}$	$-f(t)\frac{\theta S(t)+(1-\theta S(t))\log(1-\theta S(t))}{S^2(t)(1-\theta S(t))\log(1-\theta)}$	$\frac{-\theta}{\log(1-\theta)}$

3 Bayesian partition modelling

In this section we present the BPM as proposed by Holmes *et al.* (1999, 2005). The partition models are methods that split some domain of interest $\boldsymbol{\mathcal{X}} \subset \mathbb{R}^p$ $(p \geq 1)$ in disjoint regions, and assign the same probability distribution for the response variable Y in each region of $\boldsymbol{\mathcal{X}}$. In this context, the BPM partitioning $\boldsymbol{\mathcal{X}}$ by a tessellation of a structure $\mathbf{T}$ defining regions $R_m \subseteq \boldsymbol{\mathcal{X}}$, $m = 1, \ldots, M$.

One characteristic of the BPM is that assigning conjugate priors within the disjoint regions, the marginal likelihood is available for any tessellation structure. The availability of the marginal likelihood function for the tessellation structure greatly reduces the space of the models as well as the dimension of the parameter space.

In this paper we consider orthogonal hyperplanes tessellation, the hyperplanes are defined by split points $\boldsymbol{c}_{j^*}, j^* = 1, \ldots, p$. So a tessellation structure is given by $\mathbf{T} = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_p)$.

3.1 Likelihood function

The orthogonal hyperplane tessellation defines M regions in $\boldsymbol{\mathcal{X}}$, denote by $R_1, \ldots, R_M$ the regions in $\boldsymbol{\mathcal{X}}$. Let N_{mj} be the number of latent causes of the event of interest for the j-observation with power series distribution with parameter θ_m, $j = 1, \ldots, n_m$ in the region R_m.

Given N_{mj}, let $Z_{mj}^1, \ldots, Z_{mj}^{N_{mj}}$ be times of occurrence of the event of interest with cumulative distribution function $F(\cdot|\boldsymbol{\gamma}) = 1 - S(\cdot|\boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is the vector of parameters.

Let T_{mj} be as in (3) and C_{mj} the censoring time. We observe $Y_{mj} = \min\{T_{mj}, C_{mj}\}$ and δ_{mj} be the censoring indicator with $\delta_{mj} = 1$ if $Y_{mj} = T_{mj}$ and $\delta_{mj} = 0$ otherwise.

Then, the likelihood function for the complete data under uninformative censoring given the tessellation $\mathbf{T}$ is

$$L(\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}|\boldsymbol{N}, \boldsymbol{y}, \boldsymbol{\delta}) = \prod_{m=1}^{M} \prod_{j=1}^{n_m} \{S(y_{mj}|\boldsymbol{\gamma})\}^{N_{mj}-\delta_{mj}} \{N_{mj} f(y_{mj}|\boldsymbol{\gamma})\}^{\delta_{mj}} p(N_{mj}|\theta_m), \tag{7}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)^\top$ and $\boldsymbol{N} = (N_1, \ldots, N_n)^\top$ is a vector of latent variables. Note that in each region R_m the number of causes for the event of interest N_{mj} has the same probability distribution (e.g. Poisson).

We assume a Weibull distribution with vector of parameters $\boldsymbol{\gamma} = (\alpha, \lambda)^\top$ for the event time Z_{mj}. The cumulative distribution is given by

$$F(y|\boldsymbol{\gamma}) = 1 - \exp(-y^\alpha e^\lambda), \tag{8}$$

where $\alpha > 0$ and $\lambda \in \mathbb{R}$. The likelihood function given in (7) can be written as

$$L(\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}|\boldsymbol{N}, \boldsymbol{y}, \boldsymbol{\delta}) = \prod_{m=1}^{M} \exp\left(-e^\lambda \sum_{j=1}^{n_m} y_{mj}^\alpha N_{mj}\right) \prod_{j=1}^{n_m} \left(N_{mj} \alpha e^\lambda y_{mj}^{\alpha-1}\right)^{\delta_{mj}} p(N_{mj}|\theta_m).$$

3.2 Prior and Posterior distribution

According to the BPM methodology, the joint prior distribution for $(\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T})$ is given by

$$p(\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}) = p(\boldsymbol{\gamma}) p(\boldsymbol{\theta}, \mathbf{T}) = p(\boldsymbol{\gamma}) p(\boldsymbol{\theta}|\mathbf{T}) p(\mathbf{T}), \tag{9}$$

and we assume that the parameters of the Weibull distribution being independent, so $p(\boldsymbol{\gamma}) = p(\alpha)p(\lambda)$ where $\alpha \sim$ Gamma$(\mu_\alpha, \sigma_\alpha)$ and $\lambda \sim$ N$(\mu_\lambda, \sigma_\lambda^2)$, where μ_α, σ_α, μ_λ and σ_λ are hyperparameters. Considering that the parameters betwen regions of $\boldsymbol{\mathcal{X}}$ are independent, we have

$$p(\boldsymbol{\theta}|\mathbf{T}) = \prod_{m=1}^{M} p(\theta_m|\mathbf{T}). \tag{10}$$

Depending on the series function, $\eta(\theta)$ have different distributions and therefore the prior distributions for θ are also different. So, if N follow the binomial, negative binomial and logarithmic distribution then the prior distribution is the beta distribution. For the Poisson distribution the prior conjugate is the gamma distribution.

In the BPM, the latent vector $\boldsymbol{N}$ is introduced to draw samples from the posterior distribution $p(\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}|\boldsymbol{y}, \boldsymbol{\delta})$. So the joint posterior distribution for

$(\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}, \boldsymbol{N})$ is given by

$$p(\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}, \boldsymbol{N}|\boldsymbol{y}, \boldsymbol{\delta}) \propto \prod_{m=1}^{M} \exp\left\{ e^{\lambda} \sum_{j=1}^{n_m} y_{mj}^{\alpha} N_{mj} \right\} \prod_{j=1}^{n_m} \left(N_{mj} \alpha e^{\lambda} y_{mj}^{\alpha-1}\right)^{\delta_{mj}} p(N_{mj}|\theta_m) \times p(\boldsymbol{\gamma}) p(\boldsymbol{\theta}, \mathbf{T}). \tag{11}$$

The analysis analytic or numerical of posterior distribution $(\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}, \boldsymbol{N}|\boldsymbol{y}, \boldsymbol{\delta})$ is intractable in this situation. So, our aim is to simulate samples from the joint posterior distribution and for this purpose we use the method MCMC (see Brooks *et al.*, 2011).

Therefore, we need to sample from the full conditional distributions $(\boldsymbol{\theta}, \mathbf{T}|\boldsymbol{N}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{\delta})$, $(\boldsymbol{N}|\boldsymbol{\theta}, \mathbf{T}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{\delta})$ and $(\boldsymbol{\gamma}|\boldsymbol{\theta}, \mathbf{T}, \boldsymbol{N}, \boldsymbol{y}, \boldsymbol{\delta})$. So, note that to sample from $(\boldsymbol{\theta}, \mathbf{T}|\boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{\delta})$ we consider the full conditional distributions given by

$$p(\boldsymbol{\theta}, \mathbf{T}|\boldsymbol{N}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{\delta}) = p(\mathbf{T}|\boldsymbol{N}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{\delta}) p(\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{N}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{\delta}). \tag{12}$$

In this context, the BPM supposes the parameters between each region of $\boldsymbol{\mathcal{X}}$ are independent and taking into account that the full conditional distribution for $(\mathbf{T}|\boldsymbol{N}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{\delta})$ is given by

$$p(\mathbf{T}|\boldsymbol{N}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{\delta}) \propto \int p(\boldsymbol{N}|\boldsymbol{\theta}, \mathbf{T}) p(\boldsymbol{\theta}|\mathbf{T}) p(\mathbf{T}) \mathrm{d}\boldsymbol{\theta} = p(\boldsymbol{N}|\mathbf{T}) p(\mathbf{T}). \tag{13}$$

The conditional distribution of $(\mathbf{T}|\boldsymbol{N}, \boldsymbol{\gamma}, \boldsymbol{y}, \boldsymbol{\delta})$ given in (13) is nonstandard and so we use a Metropolis-Hastings MCMC scheme to sample of full conditional distribution of $\mathbf{T}$(see section 3.3). So, the analytical form of $p(\boldsymbol{N}|\mathbf{T})$ and full conditionals for each case is given by

(i) Binomial

If $N \sim \text{Bi}(K, \theta^*)$, $\theta^* \in (0, 1)$ then the conjugate prior for each θ_m^* is the beta distribution $\theta_m^* \sim \text{Beta}(a_0, a_1)$, where a_0 and a_1 are hyperparameters. Hence, the prior on $\boldsymbol{\theta}^*$ is $p(\boldsymbol{\theta}^*|\mathbf{T}) = \prod_{m=1}^{M} p(\theta_m^*|a_0, a_1)$. Therefore

$$p(\boldsymbol{N}|\mathbf{T}) = \prod_{i=1}^{n} \binom{K}{N_i} \prod_{m=1}^{M} \frac{\mathcal{B}(\sum_{j=1}^{n_m} N_{mj} + a_0, K n_m - \sum_{j=1}^{n_m} N_{mj} + a_1)}{\mathcal{B}(a_0, a_1)}, \tag{14}$$

where $\mathcal{B}(.)$ is the beta function. On the other hand, the full conditional distribution for $\theta_m^*$and N_{mj} are given respectively by

$$\theta_m^*|\boldsymbol{N}, \mathbf{T} \sim \text{Beta}\left(\sum_{j=1}^{n_m} N_{mj} + a_0, n_m - \sum_{j=1}^{n_m} N_{mj} + a_1 \right), \tag{15}$$

and

$$N_{mj}|\mathbf{T}, \boldsymbol{y}, \boldsymbol{\delta} \sim \text{Bi}\left(K - \delta_{mj}, \frac{\theta_m^* S(y_{mj}|\gamma)}{\theta_m^* S(y_{mj}|\gamma) + 1 - \theta_m^*}\right) + \delta_{mj}. \tag{16}$$

(ii) Poisson
In case that $N \sim \text{Poi}(\theta),\ \theta > 0$, then a conjugate prior for θ_m is the gamma distribution $\theta_m \sim \text{Gamma}(b_0, b_1)$, where b_0 e b_1 are two specified hyperparameters. Thus, the prior on $\boldsymbol{\theta}$ is $p(\boldsymbol{\theta}|\mathbf{T}) = \prod_{m=1}^{M} p(\theta_m|b_0, b_1)$ and hence $p(\boldsymbol{N}|\mathbf{T})$ is given by

$$p(\boldsymbol{N}|\mathbf{T}) = \prod_{m=1}^{M} \frac{1}{\prod_{j=1}^{n_m} N_{mj}!} \frac{b_1^{b_0}}{\Gamma(b_1)} \frac{\Gamma(\sum_{j=1}^{n_m} N_{mj} + b_0)}{(n_m + b_1)^{\sum_{j=1}^{n_m} N_{mj} + b_0}}, \tag{17}$$

where $\Gamma(.)$ is gamma function. On the other hand, the full conditional distribution for $\theta_m$and N_{mj} are given respectively by

$$\theta_m|\boldsymbol{N}, \mathbf{T} \sim \text{Gamma}\left(\sum_{j=1}^{n_m} N_{mj} + b_0, n_m + b_1\right), \tag{18}$$

and

$$N_{mj}|\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}, \boldsymbol{y}, \boldsymbol{\delta} \sim \text{Poi}\left(\theta_m S(y_{mj}|\gamma)\right) + \delta_{mj}. \tag{19}$$

(iii) Negative binomial
If $N \sim \text{Bn}(\tau, \theta),\ 0 < \theta < 1$, the conjugate prior for θ_m is the beta distribution $\theta_m \sim \text{Beta}(c_0, c_1)$, where c_0 and c_1 are two specified hyperparameters and so $p(\boldsymbol{\theta}|\mathbf{T}) = \prod_{m=1}^{M} p(\theta_m|c_0, c_1)$. Therefore $p(\boldsymbol{N}|\mathbf{T})$ is given by

$$p(\boldsymbol{N}|\mathbf{T}) = \prod_{i=1}^{n} \binom{\tau + N_i - 1}{\tau - 1} \prod_{m=1}^{M} \frac{\mathcal{B}(\tau n_m + c_0, \sum_{j=1}^{n_m} N_{mj} + c_1)}{\mathcal{B}(c_0, c_1)}. \tag{20}$$

The full conditional distribution for θ_m and N_{mj} are given respectively by

$$\theta_m|\mathbf{T}, \boldsymbol{N} \sim \text{Beta}\left(\sum_{j=1}^{n_m} N_{mj} + c_0, \tau n_m + c_1\right), \tag{21}$$

and

$$N_{mj}|\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}, \boldsymbol{y}, \boldsymbol{\delta} \sim \text{Bn}\left(\tau + \delta_{mj}, \theta_m \exp\left(-e^{\lambda} y_{mj}^{\alpha}\right)\right) + \delta_{mj}. \tag{22}$$

(iv) Logarithmic
In case that $N \sim \text{Lg}(\theta),\ 0 < \theta < 1$, we assign the beta distribution as prior distribution for θ_m, $\theta_m \sim \text{Beta}(d_0, d_1)$, where d_0 and d_1 are two specified hyperparameters and hence $p(\boldsymbol{\theta}|\mathbf{T}) = \prod_{m=1}^{M} p(\theta_m|d_0, d_1)$. The integral

$p(\boldsymbol{N}|\mathbf{T}) = \int p(\boldsymbol{N}|\boldsymbol{\theta}, \mathbf{T})p(\boldsymbol{\theta}|\mathbf{T})\mathrm{d}\boldsymbol{\theta}$, is not explicitly available, and thus we use numerical integration.

The full conditional for θ_m is given by

$$p(\theta_m|\boldsymbol{N}, \mathbf{T}) \propto \frac{\theta_m^{n_m+N_{mj}+d_0-1}(1-\theta_m)^{d_1-1}}{[-\log(1-\theta_m)]^{n_m}}. \tag{23}$$

Considering the assumption that N_{mj}'s are independent random variables then full conditional for N_{mj} in each region R_m is given by

$$p(N_{mj}|\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}, \boldsymbol{y}, \boldsymbol{\delta}) \propto \exp\left\{-e^{\lambda}y_{mj}^{\alpha}N_{mj}\right\} N_{mj}^{\delta_{mj}} \frac{\theta_m^{N_{mj}}}{N_{mj}+1}. \tag{24}$$

If $\delta_{mj} = 0$ the conditional distribution for number of competing causes that can produce the event of interest is

$$N_{mj}|\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}, \boldsymbol{y}, \boldsymbol{\delta} \sim \mathrm{Lg}(\theta_m S(y_{mj}|\boldsymbol{\gamma})). \tag{25}$$

However if $\delta_{mj} = 1$ then

$$p(N_{mj}|\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{T}, \boldsymbol{y}, \boldsymbol{\delta}) \propto \frac{N_{mj}}{N_{mj}+1}\left[\theta_m S(y_{mj}|\boldsymbol{\gamma})\right]^N. \tag{26}$$

For generation of N_{mj}, we considering the algorithm proposed by Kemp (1981) although adapted for the case where the logarithmic distribution is shifted.

Finally, the full conditional distributions for the parameters of the Weibull distribution $\boldsymbol{\gamma} = (\alpha, \lambda)^{\top}$ are respectively,

$$p(\lambda|\alpha, \boldsymbol{N}, \mathbf{T}, \boldsymbol{y}, \boldsymbol{\delta}) \propto e^{d\lambda}\exp\left(-e^{\lambda}\sum_{i=1}^{n}N_i y_i^{\alpha}\right)\exp\left(-\frac{(\lambda-\mu_\lambda)^2}{2\sigma_\lambda^2}\right), \tag{27}$$

$$p(\alpha|\lambda, \boldsymbol{N}, \mathbf{T}, \boldsymbol{y}, \boldsymbol{\delta}) \propto \alpha^d\left(\prod_{i=1}^{n}y_i^{\delta_i}\right)^{\alpha}\exp\left(-e^{\lambda}\sum_{i=1}^{n}N_i y_i^{\alpha}\right)p(\alpha|\mu_\alpha, \sigma_\alpha), \tag{28}$$

where $d = \sum_{i=1}^{n}\delta_i$.

3.3 Computational Strategy

The strategy computational for numerical and dichotomous predictors is given in Hoggart & Griffin (2001). Generally, categorical predictors are not necessarily dichotomous therefore the algorithm has to be modified to handle with qualitative predictors but in general form. In this context, let $X_{\mathbf{T}}$ a categorical covariate with $C_{\mathbf{T}}$ categories, $X_{\mathbf{T}} \in \{1, 2, \ldots, C_{\mathbf{T}}\}$, and suppose that ρ is a partition of $X_{\mathbf{T}}$ and M_ρ the number of cluster(subsets) for partition ρ of $X_{\mathbf{T}}$. The partition ρ is unknown, we need assign to it a

prior probability $p(\rho)$. We assume that $p(\rho)$ is a discrete uniform distribution on $\{1, \ldots, n_\rho\}$ where n_ρ is the number different partitions of $X_{\mathbf{T}}$.

Therefore, a natural relationship among the number the total numbers n_ρ of partitions and scale of $X_{\mathbf{T}}$. In the case that $X_{\mathbf{T}}$ has a nominal scale the number n_ρ is much greater than when $X_{\mathbf{T}}$ is ordinal scale. In Table 3 shows the number M_ρ of subsets and total partitions considering $C_{\mathbf{T}} = 4$, note that if $X_{\mathbf{T}}$ is ordinal variable the numbers of subsets in each partition ρ(ordered) is less if $X_{\mathbf{T}}$ out nominal variable.

Table 3: Number of clusters and total partitions for $X_{\mathbf{T}}$ if $C_{\mathbf{T}} = 4$

M_ρ	ρ (unordered)	ρ (ordered)
1	1	1
2	7	3
3	6	3
4	1	1
n_ρ	15	8

So, we denote by $\mathcal{I}$ the index set of predictor variables $\mathcal{I} = \{1, \ldots, p\}$, and $\mathcal{I}_{\mathbf{T}}$ is the index set of predictor variables present in the tessellation $\mathbf{T}$. Considering that $M = 1$ (i.e., $\mathcal{I}_{\mathbf{T}} = \emptyset$) then starting the algorithm with initialize the tessellation structure, $\mathbf{T}$, with just one randomly drawn predictor variable and choose a split point. In each iteration of the algorithm and when $1 < M < n$, we try with probability $1/3$, the first three moves. The first two moves concern the selection of covariate. The last three moves involve the categorical predictor with more than two categories.

- **Add.** A new partition can be added to the tessellation structure $\mathbf{T}$ by choosing a new splitting point of a predictor variable in $\mathcal{I}$. The splitting point is selected from the empirical distribution of the predictor variable chosen.

- **Delete.** A hyperplane can be eliminated by choosing a random predictor variable r^* present in the tessellation, $r^* \in \mathcal{I}_{\mathbf{T}}$.

- **Move.** A hyperplane can be changed by selecting a new splitting point of the empirical distribution of the selected variable in $\mathcal{I}_{\mathbf{T}}$.

- **Merge.**The number of clusters in the covariate $X_{\mathbf{T}}$ is decreased, by merging two subsets.

- **Split.**The number of clusters in the covariate $X_{\mathbf{T}}$ is increased, by splitting up one subset into two new subsets.

- **Alter.** The partition ρ for $X_{\mathbf{T}}$ is altered, but the number of subsets being equal.

The new tessellation $\mathbf{T}'$ proposal is accepted with probability:

$$\alpha(\mathbf{T}', \mathbf{T}) = \min\left\{1, \frac{p(\boldsymbol{N}|\mathbf{T}')p(\mathbf{T}')}{p(\boldsymbol{N}|\mathbf{T})p(\mathbf{T})}\right\}. \tag{29}$$

Hoggart & Griffin (2001) proposed the first three moves. In this paper, added the last 3 moves, which only concern qualitative covariate and are the main novelty of our approach. For easy explanation of proposed moves for categorical covariate, consider $X_{\mathbf{T}}$ with 4 categories, $X_{\mathbf{T}} \in \{1, 2, 3, 4\}$. Let ρ a partition of $X_{\mathbf{T}}$ with $M_\rho = 3$, $\rho = \{\{1, 3\}, \{2\}, \{4\}\}$. Therefore the number of clusters for $X_{\mathbf{T}}$ varies between 1 and 4 subsets.

Considering the move merge then two subsets in the ρ are combined, where a choice of groups is random, so have $\binom{M_\rho}{2}$ possibilities for this choice, each with equal probability $\frac{2}{M_\rho(M_\rho-1)}$. For example, we choose 2 and 4 for merged, this lead to a new partition ρ', $\rho' = \{\{1, 3\}, \{2, 4\}\}$ with $M_{\rho'} = 2$.

In this context, the move split in the algorithm, we would be splitting one subset in two. The choice is made randomly restricted to subsets with cardinality greater than one. So, let $\widetilde{M}_\rho$ the number of subsets with the number of subsets with more than one category in the partition ρ. The current partition ρ we have $\widetilde{M}_\rho = 2$ and assuming that we choose $\{1, 3\}$ then the new partition is given by $\rho = \{\{1\}, \{3\}, \{2, 4\}\}$ with $M_\rho = 3$ subsets.

For the move alter, assuming that the partition of $X_{\mathbf{T}}$ is ρ then changing the setting of subsets in ρ but the number of clusters being equal. In the current grouping ρ, the number of subsets is $M_\rho = 3$, then consider another partitions for $X_{\mathbf{T}}$ for instance $\{\{1, 4\}, \{2\}, \{3\}\}$. In the table 3, observed that if $M_\rho = 3$ we have 6 different partitions for $X_{\mathbf{T}}$(nominal).

If the categorical variable has more than four categories this algorithm can be applied although the computational cost is high. If we consider only partitions which maintain the order the number of partition of $X_{\mathbf{T}}$ decreases.

3.4 Models comparison criteria

To asses the goodness of fit of the models, we use the logarithm of pseudo-marginal likelihood (LPML) given in Ibrahim *et al.* (2001b, chap. 6). LPML is a well-established Bayesian model comparison criterion based on the conditional predictive ordinate (CPO) statistics, which is particularly suitable for the cure rate models.

Let D^{-i} denote the data with the ith observation deleted. For each model, for an observed time to event $(\delta_i = 1)$ we define $g(y_i|\boldsymbol{\vartheta}) = f_{pop}(y_i|\boldsymbol{\vartheta})$ and, for a censored time $(\delta_i = 0)$, $g(y_i|\boldsymbol{\vartheta}) = S_{pop}(y_i|\boldsymbol{\vartheta})$ where $\boldsymbol{\vartheta} = (\boldsymbol{\theta}, \boldsymbol{\gamma})^\top$.

We denote the posterior density of $\boldsymbol{\vartheta}$ given D^{-i} by $p(\boldsymbol{\vartheta}|D^{-i})$, $i = 1, \ldots, n$, therefore for the ith observation, CPO_i can be written as

$$\mathrm{CPO}_i = \int g(y_i|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta}|D^{-i})\mathrm{d}\boldsymbol{\vartheta} = \left\{\int \frac{p(\boldsymbol{\vartheta}|D)}{g(y_i|\boldsymbol{\vartheta})}\mathrm{d}\boldsymbol{\vartheta}\right\}^{-1}. \tag{30}$$

A Monte Carlo approximation of CPO_i (Chen *et al.*, 2000) is given by

$$\widehat{\mathrm{CPO}_i} = \left\{ \frac{1}{B} \sum_{b=1}^{B} \frac{1}{g(y_i|\boldsymbol{\vartheta}_b)} \right\}^{-1}, \tag{31}$$

where B denote the size of sample MCMC after the burning. Based on CPO_i's another criterion for comparison is the statistic defined by $\mathrm{LPML} = \sum_{b=1}^{B} \log(\mathrm{CPO}_i)$. The larger the LPML, the better the fit of a given model.

4 Application

The data set for illustrating our methodology was extracted from a melanoma study (the melanoma is a type of malignant cancer). A objective of the study is to evaluate the effectiveness of applying a high dosage of interferon alfa-2b as a way to prevent the recurrence of cancer. Patients were included in the study between 1991 to 1995, and follow-up was conducted until 1998. The response variable Y represents the time from patient to death or time of censoring. The original sample comprises 427 patients, 10 of whom were removed from analysis, since their tumor thickness data are missing. Therefore we have $n = 417$ patients, with 56% of censored observations. The variables include y: time (in years); x_1: treatment (0: observation, $n = 204$; 1: interferon, $n = 213$); x_2: age(in years); x_3: nodule category(1,$n = 82$; 2,$n = 87$; 3,$n = 137$; 4,$n = 111$); x_4: sex (0: male,$n = 263$; 1: female, $n = 154$); x_5: performance status(PS) means patient's functional capacity as regards his/her daily activities (0: fully active,$n = 363$ 1: other,$n = 54$) and x_6: tumor thickness (in mm). For more details related to the melanoma data Kirkwood *et al.* (2000) and Ibrahim *et al.* (2001b) may be consulted.

We consider as hyperparameters $\mu_\alpha = \sigma_\alpha = 0.1$ for the gamma distribution of the parameter α and the normal distribution with mean $\mu_\lambda = 0$ and variance $\sigma^2_\lambda = 100$ for the parameter λ.

The hyperparameters for beta distribution are $a_0 = a_1 = c_0 = c_1 = d_0 = d_1 = 1$ and finally for gamma distribution we assume $b_0 = b_1 = 0.1$.

Considering these prior densities we generated two parallel independent runs of the MCMC sampler with size 700000 for each parameter, disregarding the first 300000 iterations to eliminate the effect of the initial values and, to avoid correlation problems, we considering a spacing of length 100, obtaining a sample of size 4000 in each chain. To monitor the convergence of the MCMC sampler we resorted to the methods recommended by Cowles & Carlin (1996). Consider in the first chain initial values for λ and α equal to -5 and 5 respectively in the second chain were 5 and 9, in both chains initiate the algorithm with $\boldsymbol{N} = (1, \ldots, 1)$.

In the data set x_1, x_4 and x_5 are binary variables then the division of any of these variables follows as, if it occurs will result in two groups, for

example the variable x_4 which represents sex will be divided in male and female. In the case of the covariate x_3 with four categories $x_3 \in \{1, 2, 3, 4\}$, the idea of partition of x_3 was made considering the section 3.3 except that the choice of partitions has an order.

We tried different binomial and negative binomial models by taking $K = 1, 2, 7, 10$ and $\tau = 1, 3, 7, 13$ respectively. For the binomial model, the best fit is when $K = 10$. In the negative binomial model the best fit was when $\tau = 1$ i.e. the geometric model.

Table 4 presents the probability of splitting for each of the covariates for each model. We note that the ordinal covariate x_3 is almost always in the tessellation, so the tessellation by orthogonal hyperplanes identifies that x_3 has a significant effect on the response variable. One consequence is that the splitting probability of x_3 is very close to 1.00. A minor effect on Y is observed by considering the covariates x_2 and x_6. For the other covariates, the probability of splitting is close to zero which means that the are non-informative.

Table 4: Probability of splitting for covariates for different models.

Model	x_1	x_2	x_3	x_4	x_5	x_6
Binomial ($K = 10$)	0.001	0.123	0.999	0.001	0.002	0.020
Poisson	0.018	0.307	1.000	0.025	0.024	0.102
Negative binomial ($\tau = 1$)	0.020	0.256	1.000	0.019	0.017	0.092
Logarithmic	0.010	0.133	1.000	0.012	0.009	0.131

Table 5 show a mean posterior probabilities for each of the partitions for the covariate x_3. The BPM which leads to the partition with largest posterior probability is the partition formed by $\{1, 2, 3\}$ and $\{4\}$ for the binomial, Poison and negative binomial models. Nevertheless, in the logarithmic model the partition $\{1, 2\}$ and $\{3, 4\}$ have larger posterior probability. Moreover, it is interesting to note that the partitions $\{1\}, \{2, 3\}, \{4\}$ and $\{1\}, \{2\}, \{3, 4\}$ are separated by category 2 and this way the BPM identifies a point of change around the category 2.

Table 6 gives LPML, posterior means, standard deviations (SD) and 95% highest posterior density (HPD) interval for the parameters of Weibull for all models. Also, we calculated the estimated potential scale reduction $\widehat{R}$ (Gelman & Rubin, 1992) for the parameters of Weibull distribution, which for all parameter is close to 1, indicating good convergence.

We also note from Table 6 that, based on the LPML statistics the logarithmic model is deemed as the best fitting model. Note that, the SD of posterior

Table 5: Posterior probabilities for the partitions of the covariate x_3.

Partition	Posterior probability			
	Bi$(K = 10)$	Poi	Neg bin$(\tau = 1)$	Log
$\{1, 2, 3, 4\}$	0.000	0.000	0.000	0.000
$\{1\}, \{2, 3, 4\}$	0.005	0.002	0.002	0.016
$\{1, 2\}, \{3, 4\}$	0.148	0.164	0.214	0.486
$\{1, 2, 3\}, \{4\}$	0.639	0.766	0.340	0.090
$\{1\}, \{2\}, \{3, 4\}$	0.011	0.002	0.028	0.133
$\{1\}, \{2, 3\}, \{4\}$	0.085	0.033	0.191	0.154
$\{1, 2\}, \{3\}, \{4\}$	0.091	0.032	0.170	0.074
$\{1\}, \{2\}, \{3\}, \{4\}$	0.021	0.001	0.056	0.047

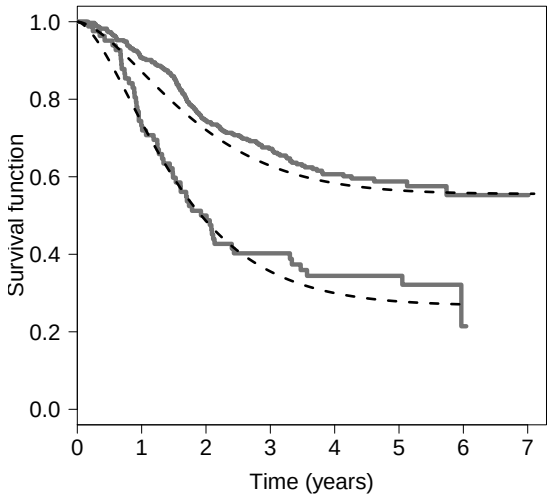
estimates of parameter λ in the binomial, Poisson and negative binomial are close, but in the logarithmic model is larger than the others models.

Table 6: Posterior summaries for the parameters of the Weibull distribution.

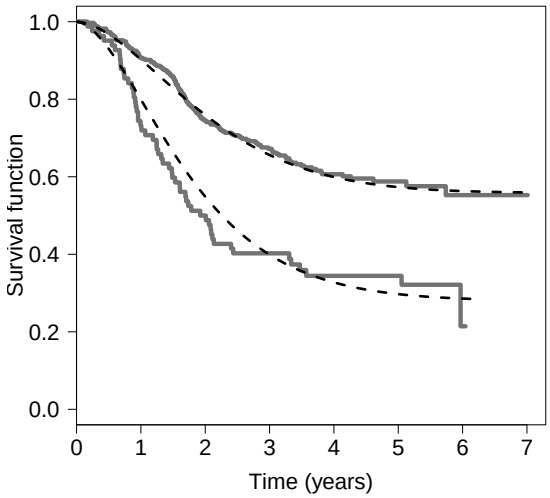
Model	LPML	Parameter	Mean	SD	95% HPD
Binomial	-521.775	α	1.599	0.109	$(1.394; 1.820)$
$(K = 10)$		λ	-1.295	0.125	$(-1.532; -1.050)$
Poisson	-521.482	α	1.721	0.116	$(1.495; 1.947)$
		λ	-1.645	0.135	$(-1.920; -1.388)$
Negative Binomial	-519.892	α	1.869	0.125	$(1.624; 2.105)$
$(\tau = 1)$		λ	-2.069	0.125	$(-2.390; -1.757)$
Logarithmic	-519.004	α	2.044	0.136	$(1.766; 2.293)$
		λ	-2.454	0.213	$(-2.890; -2.071)$

Figure 1 shows the Kaplan-Meier estimates of survival function and estimates obtained from binomial, Poisson and negative considering the partition with the most likely covariate x_3 i.e. the partition formed by $\{1, 2, 3\}$ and $\{4\}$. In fact, the binomial and Poisson models gave similar fittings but the negative binomial has a better fit.

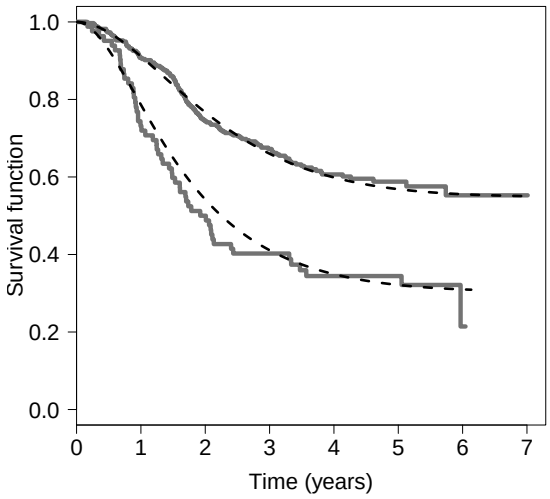
Figure 2 display the Kaplan-Meier estimates of survival function and estimates of logarithmic model considering the partition with larger posterior probability i.e. $\{1, 2\}$ and $\{3, 4\}$ for covariate x_3.



(a) Binomial($K = 10$)



(b) Poisson



(c) Negative Binomial ($\tau = 1$)

Figure 1: Kaplan-Meier curves stratified by nodule category for the clusters $\{1, 2, 3\}$ (upper curve) and $\{4\}$ (lower curve) and estimates of the survival function according to different models.

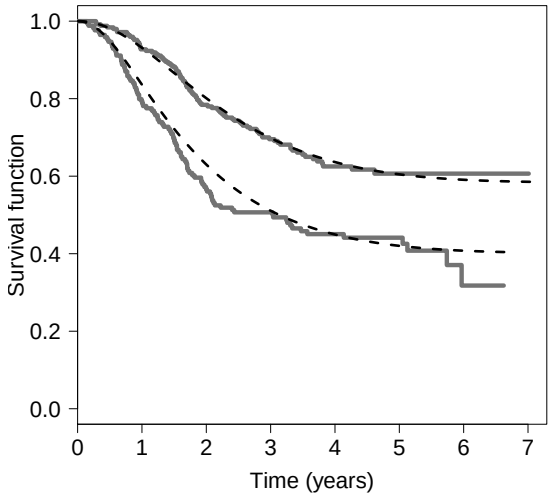


Figure 2: Kaplan-Meier curves stratified by nodule category for the clusters $\{1, 2\}$ (upper curve) and $\{3, 4\}$ (lower curve) and estimates of the survival function according to logarithmic model .

5 Discussion

In this paper, we proposed the power series cure rate based in the Bayesian partition modelling. The model proposed is a extension nonparametric for the mixture cure rate model (Berkson & Gage, 1952) and the promotion time model (Chen *et al.*, 1999).

We propose a strategy computational that considers quantitative, dichotomous and also qualitative covariates with more two categories and order. Thus, the methodology proposed extend the model proposed by Hoggart & Griffin (2001).

An important feature in Bayesian partition modeling with orthogonal hyperplanes is the natural selection of covariates. Each hyperplane divides the data set in only one covariate and thus the hyperplanes are included when the covariate affects the fit of the model.

According to our computational strategy, the partition of ordinal covariates is performed respecting an order, and this is novelty. We note that the presence of order in covariate qualitative leads to the need of adaptation of the adopted simulation procedure by the fact that the categories can not be grouped randomly.

In general, the Bayesian partition model with orthogonal hyperplanes proved to be efficient on a large data set and when using the power series cure rate model. Moreover, in survival studies, time-dependent covariates may be available. The tessellation with the computational strategy given by

hyperplanes as in Section 3.3 has not been researched yet. Research in this direction can be seen as a step further to generalize the proposed framework to include time-dependent covariates.

Acknowledgements

The research was partially supported by the Brazilian Organizations FAPESP, CNPq and CAPES

References

Berkson, J. & Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**, 501–515. 2, 16

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, **11**, 15–53. 2

Brooks, S., Gelman, A., Jones, G. L. & Meng, X.-L., editors (2011). *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC, Boca Raton, FL. 7

Cancho, V. G., Rodrigues, J. & de Castro, M. (2011). A flexible model for survival data with a cure rate: a Bayesian approach . *Journal of Applied Statistics*, **38**, 57 – 70. 2

Chen, M.-H., Ibrahim, J. G. & Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**, 909–919. 2, 16

Chen, M.-H., Shao, Q. M. & Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer, New York. 12

Cooner, F., Banerjee, S., Carlin, B. P. & Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*, **102**, 560–572. 2

Cowles, M. K. & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, **91**, 883–904. 12

de Castro, M., Cancho, V. G. & Rodrigues, J. (2009). A Bayesian long-term survival model parametrized in the cured fraction. *Biometrical Journal*, **51**, 443–55. 2

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–472. 13

Green, P. J. & Sibson, R. (1978). Computing Dirichlet tessellations in the plane. *Computer Journal*, **21**, 168–173. 2

Gu, Y., Sinha, D. & Banerjee, S. (2011). Analysis of cure rate survival data under proportional odds model. *Lifetime Data Analysis*, **17**, 123–134. 2

Heikkinen, J. (1998). Curve and surface estimation using dynamic step functions. In D.Dey, editor, *Practical Nonparametric and Semiparametric Bayesian Statistics, no. 133 in Lecture Notes in Statistics, chap. 14*, pages 255–272, New York. Springer-Verlag. 2

Hoggart, C. & Griffin, J. E. (2001). A Bayesian partition model for customer attrition. In E. I. George, editor, *Bayesian Methods with Applications to Science, Policy, and Official Statistics(Selected Papers from ISBA 2000)*, pages 61–70, Creta,Greece. International Society for Bayesian Analysis, Proceedings of the the Sixth World Meeting of the International Society for Bayesian Analysis. 2, 9, 11, 16

Holmes, C. C., Denison, D. G. T. & Mallick, B. K. (1999). Bayesian partitioning for classification and regression. Technical report, Department of Mathematics, Imperial College. 2, 5

Holmes, C. C., Denison, D. G. T., Ray, S. & Mallick, B. K. (2005). Bayesian prediction via partitioning. *Journal of Computational and Graphical Statistics*, **14**, 811–830. 5

Ibrahim, J. G., Chen, M.-H. & Sinha, D. (2001a). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, **57**, 383–388. 3

Ibrahim, J. G., Chen, M.-H. & Sinha, D. (2001b). *Bayesian Survival Analysis.*. Springer, New York. 11, 12

Johnson, N. L., Kemp, A. W. & Kotz, S. (2005). *Univariate Discrete Distributions* . Wiley, Hoboken, third edition. 3

Kemp, A. W. (1981). Efficient generation of logarithmically distributed pseudo-random variables . *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **30**, 249–253. 9

Kim, S., Chen, M.-H., Dey, D. K. & Gamerman, D. (2007). Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Analysis*, **13**, 17–35. 3

Kirkwood, J. M., Ibrahim, J. G., Sondak, V. K., Richards, J., Flaherty, L. E., Ernstoff, M. S., Smith, T. J., Rao, U., Steele, M. & Blum, R. H. (2000). High-

and low-dose interferon alfa-2b in high-risk melanoma: First analysis of intergroup trial e1690/s9111/c9190. *Journal of Clinical Oncology*, **18**, 2444–2458. 12

Kosambi, D. D. (1949). Characteristic properties of series distributions. *Proceedings of the National Institute for Science, India*, **15**, 109–113. 3

Maller, R. A. & Zhou, S. (1996). *Survival Analysis with Long-Term Survivors.*. Wiley, New York,NY. 2

Noack, A. (1950). A class of random variables with discrete distributions. *The Annals of Mathematical Statistics*, **21**, 127–132. 3

Rodrigues, J., Cancho, V. G., de Castro, M. & Louzada-Neto, F. (2009a). On the unification of long-term survival models. *Statistics & Probability Letters*, **79**, 753–759. 4

Rodrigues, J., de Castro, M., Cancho, V. G. & Balakrishnan, N. (2009b). COM-Poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference*, **139**, 3605–3611. 2

Rodrigues, J., de Castro, M., Balakrishnan, N. & Cancho, V. G. (2011). Destructive weighted Poisson cure rate models. *Lifetime Data Analysis*, **17**, 333–346. 2

Tsodikov, A. D., Ibrahim, J. G. & Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, **98**, 1063–1078. 2, 4

Yakovlev, A. Y. & Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore. 2, 3

Yao, Y.-C. (1984). Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. *The Annals of Statistics*, **12**, 1434–1447. 2

Yin, G. & Ibrahim, J. G. (2005). Cure rate models: A unified approach. *The Canadian Journal of Statistics*, **33**, 559–570. 2